Teaching Ordinal and Criterion Scaling in Multiple Regression

Randall E. Schumacker
University of North Texas

John Delane Williams
University of North Dakota

This article illustrates the appropriate use of ordinal and criterion scaling techniques in multiple regression. Since multiple regression is a widely used data analytic technique, it is important to know how special coding is done to answer certain research questions. These coding techniques involve effect, characteristic, ordinal, or contrast coding of vectors for proper interpretation of statistical hypotheses.

The focus of this article is to demonstrate ordinal and criterion scaling techniques in multiple regression. Instructors might wish to include these topics when teaching multiple regression. Examples of each coding technique are presented to aid in understanding the approach and interpreting results.

Various authors have previously discussed specific coding strategies needed in multiple regression to answer certain research questions. For example, binary, or characteristic coding can be used to test research hypotheses involving group mean differences (Draper & Smith, 1966; Kerlinger & Pedhazur, 1973; Williams, 1974; McNeil, Kelly, & McNeil, 1975) or be used in multiple comparisons (Williams, 1974, 1976, 1980). Similarly, contrast coding can be used to investigate complex comparisons and other types of research questions (Lewis & Mouw, 1978). Effect, orthogonal, and polynomial coding techniques have also been elaborated (Cohen & Cohen, 1975; Pedhazur, 1982). Within orthogonal techniques, Helmert contrasts (see Bock, 1975) or polynomial regression can be completed. Newman (1988) presented several examples of how various coding strategies in multiple regression yield a t-test, analysis of variance, chi-square, discriminant, and other statistical results. Newman, Williams and Bobner (1982) had shown earlier that the Cochran Q test could be readily conceptualized into a regression format; they used a Monte Carlo study to show that the outcomes of using regression virtually coincided with the traditional Cochran Q analysis. Coding for two-way analysis of variance questions has also received close scrutiny (Bottenberg & Ward, 1963; Cohen, 1968; Overall & Spiegel, 1969; Overall, Spiegel, & Cohen, 1975; Speed & Hocking, 1976; Timm & Carlson, 1975; Ward & Jennings, 1973; Williams, 1972, 1977b).

Ordinal and criterion scaling techniques have not received as much attention in the research literature as binary, effect, orthogonal, and polynomial coding strategies. Their application to specific Likert (ordinal) scaled questionnaire data and repeated measures designs (criterion scaled), for example, have not been as well understood. Consequently, this paper presents ordinal and criterion scaling techniques in multiple regression.

Ordinal Scaling

Boyle (1970), Lyons and Carter (1971), and Lyons (1971) have elaborated on the use of ordinal scaling in multiple regression. Basically, ordinal scaling permits the interpretation of Likert (ordinal scaled) questionnaire data using multiple regression techniques. This approach defines the regression line between each ordinal point individually, disregarding the linear least squares rule applied to the entire set of data across the scale. The technique applies an eta-squared function and the relative contribution made by each segment of the ordinal variable; in essence, computing the slope of each regression line connecting the Y-means for successive categories of the ordinal scaled variable.

The cumulative nature of the coding in the regression equation is the basis for interpretation of the ordinal coefficients. Consequently, each successive beta weight represents the change in predicted Y from the previous category of the ordinal variable to the next. The ordinal approach doesn't force a uniform b_{yx} for the full range of values, but instead allows a separate prediction for each interval (a separate b_i for each segment between levels of the ordinal variable), and thus a maximum nonlinear prediction of Y given a specific category of the ordinal variable. The non-linear eta-squared value therefore will be equal to or greater than the linear least squares R-squared value. The ordinal interpretation is found in the b_i values themselves which are additive across categories of the ordinal variable.

Figure 1 Computer Program and Output: Ordinal Coding.

PROGRAM

```
TITLE     Regression analysis using ordinal coded variable
COMMENT   Interpretation of regression weights is additive
DATA LIST FIXED RECORDS=1/ Y 1-2 SA 4 A 6 D 8
VARIABLE LABELS Y 'score' SA 'strongly agree' A 'agree' D 'disagree'
BEGIN DATA
10  1  1  1
14  1  1  1
13  1  1  1
11  1  1  1
 9  0  1  1
11  0  1  1
12  0  1  1
12  0  1  1
 6  0  0  1
 9  0  0  1
10  0  0  1
11  0  0  1
 6  0  0  0
11  0  0  0
 7  0  0  0
 8  0  0  0
END DATA
REGRESSION VARIABLES = Y SA A D/
NOORIGIN/
DEPENDENT= Y/
METHOD= ENTER SA A D
FINISH
```

COMPUTER OUTPUT

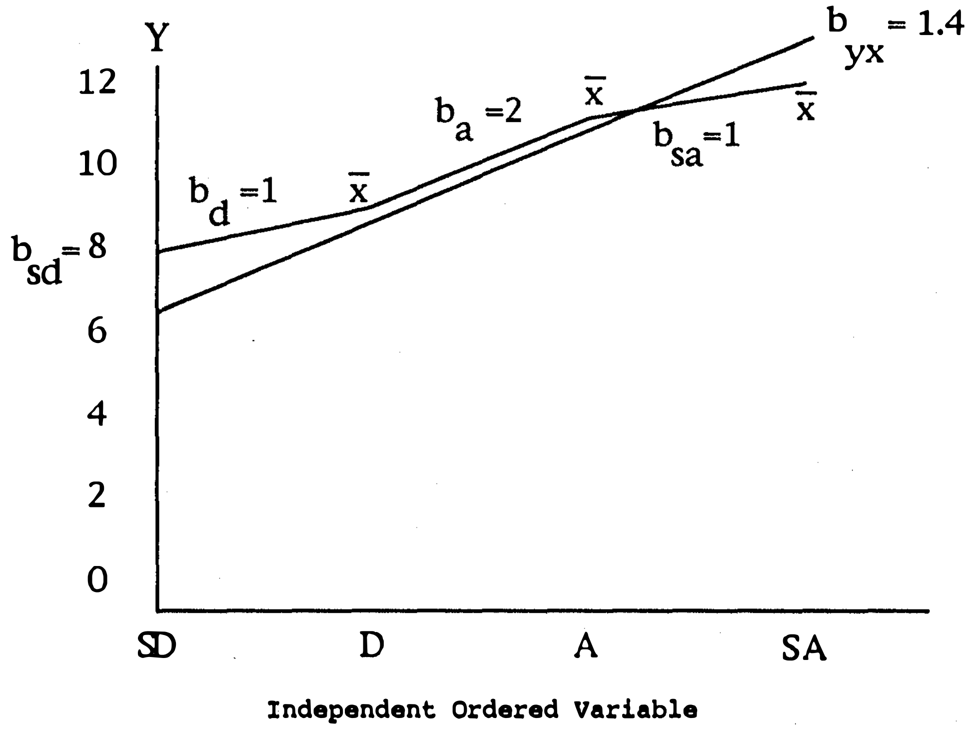
1. Analysis of Variance Summary Table

Source	df	Sum of Squares	Mean Square	F Value	Prob>F	R-Square
Model	3	40.00	13.33	3.63	.04	.47619
Error	12	44.00	3.67			
Total	15	84.00				

2. Parameter Estimates

Variable	df	Parameter Estimate	Standard Error	T for H0:	Prob > ITI
INTERCEPT (SD)	1	8.00	0.95	8.35	.0001
SA	1	1.00	1.35	0.74	.47
A	1	2.00	1.35	1.48	.16
D	1	1.00	1.35	0.74	.47

Figure 2 b_i Values for Ordinal Scaled Mean Compared to the Common Slope Value b_{yx}.



An example computer program with the required ordinal coding and the regression analysis output are in Figure 1. Corresponding parameter estimates are graphed in Figure 2, where the common slope, b_{yx}, is drawn relative to the successive b_i ordinal parameter estimate values. The dependent variable is score and the k-1 vectors represent responses to an ordinally scaled question. The subject responses are coded across k-1 vectors with the neutral or undecided response omitted. In the example, strongly disagree (SD), the first level of the ordinally scaled variable, has been omitted. Consequently, the value for b_{sd} is computed as the intercept term and reflects the starting value for interpreting change in the predicted Y values for successive ordinal categories. The ordinal coding causes the ordered variable means to range from lowest to highest.

Typically, the means for each ordinal category will deviate from linearity such that predicting Y would occur to a lesser extent than would be possible using a non-linear function with line segments between the ordinal categories. The inherent feature in interpreting ordinal coding is the cumulative or aggregate nature of responses across the categories. The predicted Y values (Y') are the respective sums of all the b_i values plus the intercept value (b_i for the omitted vector, e.g., $b_{sd} = 8$, and $b_d = 1$, $b_a = 2$, $b_{sa} = 1$ for each successive b_i). Each successive b_i reflects the change in the predicted Y value from the previous category. The parameter estimates are also additive to produce the next ordered variable group mean on Y, e.g., $\bar{X}_{sd} = 8$, $\bar{X}_{d} = 9$, $\bar{X}_{a} = 11$, $\bar{X}_{sa} = 12$. The important points for illustrative purposes is simply that ordinal coding does not compute a common slope (b_{yx}) across the full range of Y values, but instead permits separate prediction (separate b_i) between each successive ordinal category, and that a cumulative effect is apparent across levels of ordered categories.

How does ordinal scaling compare to other techniques? It might be seen as interesting that ordinal scaling yields similar results to binary coding the separate responses. The Likert response is binary coded for k-1 responses, (i.e., SA is binary coded as 1 if SA, 0 otherwise; A is binary coded as 1 if A, 0 otherwise; D is binary codes as 1 if D, 0 otherwise). Then, remarkably, if these three variables are used as predictors, an identical analysis of variance summary table to that in Figure 1 is formed. For both the ordinal scaling and the binary coding, R^2 =.47619. Performing the binary coding rather than assuming interval level data was suggested by McNeil and Kelly (1970). The two coding approaches are similar because the process of creating the ordinal variables is very similar to testing for the departure from linearity as described by Bottenberg and Ward (1963).

The advantage of ordinal scaling is its intuitive appeal; it has a similarity to the unfolding technique described by Coombs (1964). To test for the departure from linearity, one additional vector (predictor) is needed; it is coded 4 for SA, 3 for A, 2 for D, and 1 for SD. The regression analysis yields a R^2 =.46667, b_{yx} = 1.4, and intercept = 6.5 (see Figure 2). The F-test for departure from linearity is given by:

$$F = \frac{\left(R^2_{FULL} - R^2_{RESTRICTED}\right)/(k-2)}{\left(1-R^2_{FULL}\right)/(N-k)} \quad [1]$$

where R^2_{FULL} =.47619, calculated using either the ordinal scaling or the binary coding; where $R^2_{RESTRICTED}$ =.46667, calculated using the single predictor; k=4; and where N=16. Using these values, F = .213 which is not significant. This indicates that the ordered variable means do not significantly deviate from linearity.

To summarize, a variable's linear effect on a dependent variable can only be less than or equal to the nonlinear effect. A straight regression line can poorly describe the relationship between means to the extent that the means do not lie on a straight line (curvilinear path of categorical means). The test of departure from linearity will assess whether this exists. Moreover, to include ordinal variables in regression analysis is appealing, especially when one can discover which intervals, if any, contribute more or less and the degree of change from one category to another.

Criterion Scaling

Criterion scaling was first developed by Beaton (1969a; 1969b) to solve certain problems encountered in multiple regression. A basic problem occurred when using categorical variables because N - 1 vectors had to be created using dummy or effect coding. If the number of categorical independent predictors became large, then the number of vectors became overwhelming. A second problem pertained to variable selection methods where the categorical vectors might only be partially selected making interpretation difficult. Missing data on one of the predictor variables also presented a problem and usually meant exclusion from analysis even when a criterion value was present. These problems were resolved using the criterion scaling approach.

A categorical variable is criterion scaled when it is transformed into a single vector in which each individual score is replaced with the mean of the group to which the individual belongs (Pedhazur, 1982, p. 387). By criterion scaling a single categorical variable, the multiple regression analysis reduces to a bivariate regression analysis in which the dependent variable is regressed on the criterion scaled variable. This holds true regardless of the number of categories and for equal or unequal n's (Williams, 1977a). Comparing this to the traditional analysis of variance, this process removes within cell variability and leaves only variability due to group differences.

If the study involves several categorical variables, each variable can be criterion scaled separately and the criterion scaled variables used in the regression equation. Criterion scaling is therefore very useful when using a variable selection procedure to obtain the best set of independent predictors. For example, if five categorical variables resulted in 15-20 coded vectors, it may be difficult to have these sets of vectors added to or dropped from the equation as a set. (See however, Williams and Lindem, 1971 for a description of setwise regression). With criterion scaling, each categorical variable is represented by a single criterion scaled vector, so the problem is averted.

The degrees of freedom associated with criterion scaled variables however present a unique problem. Typically, the categorical variable is associated with k - 1 degrees of freedom. However, in criterion scaling, the variable will only have one degree of freedom reported by most computer programs. The actual degrees of freedom are k - 1. An example of criterion scaling is shown in Table 1. The three dummy coded vectors, D1, D2, and D3 are reduced to a single vector X1. This new vector is criterion scaled and contains the mean on Y for each respective dummy coded vector. The bivariate regression equation then becomes: $Y' = b_0 + b_2 X_2$, which would yield the same results as the dummy coded vectors in a regression analysis.

Table 1 Criterion Scaling Example

Score	Dummy Coded Vectors			Criterion Scale
Y	D1	D2	D3	X2
4	1	0	0	5
5	1	0	0	5
6	1	0	0	5
7	0	1	0	8
8	0	1	0	8
9	0	1	0	8
10	0	0	1	11
11	0	0	1	11
12	0	0	1	11

Criterion Scaling - Repeated Measures

Pedhazur (1977, 1982) and Williams (1977a) elaborated the usefulness of criterion scaling in treatment by subject repeated measures designs. This approach involves reducing the coding of N - 1 vectors to represent subjects into a single vector whereby each subject in a treatment group receives the sum of the criterion scores for that group. This single vector in a bivariate regression analysis yields the same R-squared value as does the N - 1 binary coded subject vectors. The traditional analysis proceeds with three linear models (Williams, 1974): (a) treatment effects, (b) subject effects, and (c) combined treatment and subject effects.

The three regression equations can be expressed as:

$$Y_{treat} = b_0 + b_2X2 + b_3X3 + e_1 \quad [2]$$

$$Y_{subj} = b_0 + b_4X4 + b_5X5 + \ldots b_{12}X12 + e_2 \quad [3]$$

$$Y_{comb} = b_0 + b_2X2 + b_3X3 + \ldots b_{12}X12 + e_3 \quad [4]$$

The criterion scaling approach also involves three linear models, however, the subject effects in equation [3] and the combined treatment and subject effects in equation [4] would be substantially reduced as follows:

$$Y_{subj} = b_0 + b_1X1 + e_2 \quad [5]$$

$$Y_{comb} = b_0 + b_1X1 + b_2X2 + b_3X3 + e_3 \quad [6]$$

The criterion scaling for this example treatment by subjects design is in Table 2. The variable, X_1, is the criterion scaled vector which reduces the N-1 subject vectors (X4 to X12) into a single vector for bivariate regression analysis to obtain the subjects effect in equation [5]. Score sums on Y rather than means were used in the criterion scaled vector, which is permissible.

Table 2 Criterion Scaling in Treatment by Subject Design[a]

Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
10	33	1	0	1	0	0	0	0	0	0	0	0
11	33	0	1	1	0	0	0	0	0	0	0	0
12	33	0	0	1	0	0	0	0	0	0	0	0
13	42	1	0	0	1	0	0	0	0	0	0	0
14	42	0	1	0	1	0	0	0	0	0	0	0
15	42	0	0	0	1	0	0	0	0	0	0	0
14	38	1	0	0	0	1	0	0	0	0	0	0
13	38	0	1	0	0	1	0	0	0	0	0	0
11	38	0	0	0	0	1	0	0	0	0	0	0
6	21	1	0	0	0	0	1	0	0	0	0	0
7	21	0	1	0	0	0	1	0	0	0	0	0
8	21	0	0	0	0	0	1	0	0	0	0	0
10	35	1	0	0	0	0	0	1	0	0	0	0
14	35	0	1	0	0	0	0	1	0	0	0	0
11	35	0	0	0	0	0	0	1	0	0	0	0
15	35	1	0	0	0	0	0	0	1	0	0	0
12	35	0	1	0	0	0	0	0	1	0	0	0
8	35	0	0	0	0	0	0	0	1	0	0	0
14	44	1	0	0	0	0	0	0	0	1	0	0
15	44	0	1	0	0	0	0	0	0	1	0	0
15	44	0	0	0	0	0	0	0	0	1	0	0
12	39	1	0	0	0	0	0	0	0	0	1	0
17	39	0	1	0	0	0	0	0	0	0	1	0
10	39	0	0	0	0	0	0	0	0	0	1	0
22	60	1	0	0	0	0	0	0	0	0	0	1
21	60	0	1	0	0	0	0	0	0	0	0	1
17	60	0	0	0	0	0	0	0	0	0	0	1
11	44	1	0	0	0	0	0	0	0	0	0	0
18	44	0	1	0	0	0	0	0	0	0	0	0
15	44	0	0	0	0	0	0	0	0	0	0	0

[a] Y (criterion), X1 (criterion scaled), X2 and X3 (Treatments) X4 to X12 are the N-1 subject coded vectors.

Table 3 Criterion Scaled Analysis of Variance Summary

Source	SS	df	MS	F	R^2
Treatments	21.67	2	10.83	2.22	.0532
Subjects	297.63	9			.7314
Residual	87.66	18	4.87		.2154
Total	406.96	29			1.0000

The results obtained are identical with the traditional N - 1 coded subject vectors (sum of squares and R-squared values), except the degrees of freedom are reported as N - 1, because X1 contains (N -1) linearly independent vectors. The researcher must correct the degrees of freedom when using general purpose multiple regression computer programs. The criterion scaled analysis of variance results are presented in Table 3. The combined subject and treatment effects yield $R^2 = .7846$.

If multiple comparisons were of interest, the tests of significance for comparing Group 1 to Group 3 and Group 2 to Group 3 would be derived from the reported t tests for the regression coefficients b_2 and b_3 respectively in equation [6] (output not shown). Because of the degrees of freedom issue already described, the reported t's would need to be adjusted by multiplying by:

$$\sqrt{\frac{N-S-g+1}{N-g-1}} \qquad [7]$$

where N is the number of observations (treatment by subject combinations), S is the number of subjects and g is the number of groups; an appropriate table, such as Dunnett's (1964), Dunn's (1961), or some other multiple comparison procedure being used now can be entered.

Summary

This article focused on presenting ordinal and criterion scaling techniques in multiple regression. The computer programs, coding, and output examples reflect the need to teach how special coding can be used to answer certain research questions. These coding techniques also highlight the need for proper interpretation of results.

Ordinal scaling techniques in multiple regression provide for the analysis and interpretation of ordinal variables such as Likert scaled questionnaire data. The regression weights provide a step interpretation between each point on the scale or the degree of change from one category to the next. A test for the departure from linearity can also be conducted.

Although only one example was presented, criterion scaling techniques can solve many of the problems encountered in multiple regression, namely, (a) the use of extensive categorical variables,* (b) the selection of the best set of predictor variables (maximize R-squared), (c) coding of subjects in repeated measures designs, and (d) the handling of predictor variables with missing values. Criterion scaled vectors may contain either means or sums in the case of equal n's. The criterion scaling technique is applicable to both linear and non-linear regression line fitting (Hinkle, Wiersma, & Jurs, 1988, pp. 540-544). Continuous predictor variables can also be criterion scaled by dividing them into equal intervals with the scores in each interval coded the value of the mean on the criterion for that category. In the case of multiple predictors, each predictor is coded into a single vector, then all possible regression techniques can be applied. Remember, however, that the degrees of freedom must be adjusted to N-1 not the reported df = 1. Mixed regression models which combine categorical and continuous predictors are also possible (Gocka, 1973).

References

Beaton, A. E. (1969a). Scaling criterion of questionnaire items. *Socio-Economic Planning Sciences, 2*, 355-362.

Beaton, A. E. (1969b). Some mathematical and empirical properties of criterion scaled variables. In G. W. Mayeske (Ed.), *A study of our nation's schools*. Washington, D.C.: Office of Education.

Bock, R. D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill.

Bottenberg, R. A., & Ward, J. H. (1963). *Applied multiple linear regression*. Lackland Air Force Base, Texas: Aerospace Medical Division, AD 413128.

Boyle, R. P. (1970). Path analysis and ordinal data. *American Journal of Sociology, 75*, 461-480.

Cohen, J. (1968). Multiple regression as a general data analytic system. *Psychological Bulletin, 81*, 335-343.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Draper, N. R., & Smith, H. (1966). *Applied regression analysis*. New York: Wiley.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*, 52-64.

Dunnett, C. W. (1964). New tables for multiple comparisons with a control. *Biometrics, 20*, 482-491.

Gocka, E. F. (1973). Stepwise regression for mixed mode predictor variables. *Educational and Psychological Measurement, 33*, 319-325.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). *Applied* *statistics for the behavioral sciences.* Boston: Houghton Mifflin.

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research.* New York: Holt, Rinehart, & Winston.

Lewis, E. L., & Mouw, J. T. (1978). *The use of contrast coefficients.* Carbondale, IL: Southern Illinois University Press.

Lyons, M. (1971). Techniques for using ordinal measures in regression and path analysis. In H. L. Costner (Ed.), *Sociological methodology* (pp. 147-171). San Francisco, CA: Josey-Bass.

Lyons, M., & Carter, T. M. (1971). Comment on Boyle's path analysis and ordinal data. *American Journal of Sociology, 76,* 1112-1132.

McNeil, K. A., & Kelly, F. J. (1970). Express functional relationships among data rather than assume "intervalness". *The Journal of Experimental Education, 39,* 43-48.

McNeil, K. A., Kelly, F. J., & McNeil, J. T. (1975). *Testing research hypotheses using multiple linear regression.* Carbondale, IL: Southern Illinois University Press.

Newman, I. (1988, October). *There is no such thing as multivariate analysis: all analyses are univariate!* Presidential address presented at the annual Midwestern Educational Research Association meeting, Chicago, IL.

Newman, I., Williams, J. D., & Bobner, R. (1982). Dependent binary data: A comparison of regression and chi-square solutions. *Proceedings of the American Statistical Association, Statistical Education,* 178-183.

Overall, J. E., & Spiegel, D. H. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin, 72,* 311-322.

Overall, J. E., Spiegel, D. H., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. *Psychological Bulletin, 82,* 182-186.

Pedhazur, E. J. (1977). Coding subjects in repeated measures designs. *Psychological Bulletin, 84,* 298-305.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research.* New York: CBS College Publishing.

Speed, F. A., & Hocking, R. R. (1976). The use of R () - notation with unbalanced data. *The American Statistician, 30,* 30-33.

Timm, N. H., & Carlson, J. E. (1975). Analysis of variance through full rank models. *Multivariate Behavioral Research: Monograph, 75*(1).

Ward, J. H., & Jennings, E. (1973). *Introduction to linear models.* Englewood Cliffs, NJ: Prentice Hall.

Williams, J. D. (1972) Two way fixed effects analysis of variance with disproportionate cell frequencies. *Multivariate Behavioral Research, 7,* 67-83.

Williams, J. D. (1974). *Regression analysis in educational research.* New York: MSS Publishing.

Williams, J. D. (1976). Multiple comparisons by multiple linear regression. Monograph Series No. 2, *Multiple Linear Regression Viewpoints, 7* (1).

Williams, J. D. (1977a). A note on coding the subjects effect in treatments by subjects designs. *Multiple Linear Regression Viewpoints, 8*(1), 32-35.

Williams, J. D. (1977b). Full rank and non-full rank models with contrast and binary coding systems for two-way disproportionate cell frequency analyses. *Multiple Linear Regression Viewpoints, 8*(1), 1-18.

Williams, J. D. (1980). Multiple comparisons in higher dimensional designs. Monograph Series No. 5, *Multiple Linear Regression Viewpoints, 10*(3).

Williams, J. D., & Lindem, A. C. (1971). Setwise regression analysis: A stepwise procedure for sets of variables. *Educational and Psychological Measurement, 31,* 747-748.

RELATED AERA SPECIAL INTEREST GROUPS

As a member of the SIG/MLR, the AERA Special Interest Groups listed below also might be of interest to you. The 1993 AERA Annual Meeting provides the opportunity to attend a session sponsored by one of these groups or to contact an officer to obtain additional information and a membership application.

SIG/Educational Statisticians
CHAIR: Schuyler Huck, University of Tennessee
SECRETARY: Lynne Edwards, University of Minnesota

SIG/Professors of Educational Research
CHAIR: William Schafer, University of Maryland
SECRETARY: Bruce Thompson, Texas A&M University

SIG/Structural Equation Modeling
CHAIR: Ralph Mueller, University of Toledo
SECRETARY: Jack Pollard, SWRL